

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»**

СУЛІМА СВІТЛАНА ВАЛЕРІЇВНА



УДК 621.391

**МЕТОДИ РЕКОНФІГУРАЦІЇ ОБЧИСЛЮВАЛЬНИХ РЕСУРСІВ БАЗОВОЇ
МЕРЕЖІ НА ОСНОВІ ТЕХНОЛОГІЇ ВІРТУАЛІЗАЦІЇ**

Спеціальність: 05.12.02 – телекомунікаційні системи та мережі

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Київ – 2019

Дисертацією є рукопис.

Робота виконана на кафедрі інформаційно-телекомунікаційних мереж Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» Міністерства освіти і науки України.

Науковий керівник: кандидат технічних наук, старший науковий співробітник
Скулиш Марія Анатоліївна
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»,
доцент кафедри інформаційно-телекомунікаційних мереж

Офіційні опоненти: доктор технічних наук, професор
Печурін Микола Капітонович
Національний авіаційний університет,
професор кафедри комп'ютерних систем та мереж

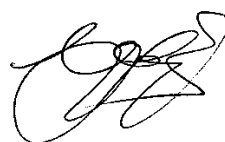
кандидат технічних наук, доцент
Гринкевич Ганна Олександрівна
Державний університет телекомунікацій,
доцент кафедри телекомунікаційних систем та мереж

Захист відбудеться «20» травня 2019 р. о 15 годині на засіданні спеціалізованої вченої ради Д 26.002.14 в Національному технічному університеті України «Київський політехнічний інститут імені Ігоря Сікорського» за адресою: 03056, м. Київ, пр. Перемоги, 37, корп. №1, ауд. 255.

З дисертацією можна ознайомитись в науково-технічній бібліотеці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» за адресою: 03056, м. Київ, пр. Перемоги, 37.

Автореферат розісланий «11» квітня 2019 р.

Учений секретар спеціалізованої
вченої ради



Л.О. Уривський

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. У мережі мобільного стільникового зв'язку останнім часом спостерігається швидкий розвиток. Сьогоднішні мобільні мережі здатні запропонувати більш високі швидкості передачі даних, з інтеграцією більшої кількості послуг і гарантією високої якості сприйняття. Тим не менш, цей розвиток означає також, що обсяг даних, які передаються в мобільній мережі, зростає, відповідно зростає обсяг службового трафіку. При цьому середня вимога сигналізації на одного абонента до 42% вище у LTE (Long-Term Evolution) порівнюючи зі стандартом зв'язку минулого покоління. Крім того, конкуренція на ринку вимагає більш швидкого розгортання послуг і еластичність в зміні критеріїв обслуговування і здатність справлятися з більш високими вимогами обслуговування. Тому виникає необхідність у керуванні службовим трафіком з метою забезпечення потрібної якості обслуговування кінцевих користувачів та належного використання ресурсів мережі оператора зв'язку.

В таких умовах оператори змушені нарощувати інфраструктуру мережі для забезпечення процесу обслуговування телекомунікаційних сервісів на заданому рівні якості. Протягом дня навантаження змінюється, відповідно до останніх досліджень до 80% обчислювальної потужності базових станцій і до половини потужності базової мережі є невикористаними. Це призводить до неефективного використання ресурсів, а також до високого рівня споживання енергії, що знижує економічну ефективність мережі для операторів мобільного зв'язку.

Поява концепції віртуалізації мережевих функцій відкриває нові можливості для світу телекомунікаційних систем, в той же час виникає необхідність у нових підходах, моделях та методах організації процесу обслуговування сервісів. Використання віртуальних серверів для вирішення задач базової мережі мобільного зв'язку може значно спростити процес організації ресурсів на сервері обслуговування, забезпечить його масштабованість та відмовостійкість.

Принцип Віртуалізації Мережевих Функцій (NFV) спрямований на перетворення мережевих архітектур шляхом впровадження мережевих функцій в програмному забезпеченні, що може працювати на стандартній апаратній платформі. Згідно зі специфікацією інституту ETSI мережева функція є функціональним блоком в межах мережевої інфраструктури, який має чітко визначені зовнішні інтерфейси і чітко визначену функціональну поведінку. Мережевими функціями є компоненти в мережі LTE EPC (Evolved Packet Core), наприклад MME, HSS, PGW, SGW, які для випадку NFV будуть розгортатись на базі системи даних центів, з використанням орендованих обчислювальних ресурсів (ядра CPU, пам'ять, дисковий простір, мережева карта), що можуть виділятися і перерозподілятися в процесі роботи в залежності від фактичних вимог навантаження.

Вітчизняні та закордонні вчені пропонують нові архітектурні рішення для впровадження концепцій віртуалізації мережевих сутностей з використанням віртуальних обчислювальних ресурсів для побудови мереж всіх рівнів в тому числі базової мобільної мережі. Зокрема у своїх роботах ці проблеми розглядають Баумгартнер А., Баста А., Глоба Л.С., Климаш М.М. та інші вчені України та світу.

Невирішеним залишається ряд проблем. Необхідно врахувати гібридність обслуговуючого середовища, де поряд із спеціалізованими апаратними засобами з обмеженими характеристиками працюють гнучкі добре масштабовані віртуальні обслуговуючі сутності розміщені у орендованих хмарних дата центрах. Тому актуальним є завдання організації обчислювальних ресурсів вузлів обслуговування та потоків між ними у гібридному середовищі, яке складається із апаратних засобів телекомунікацій та віртуальних обчислювальних сутностей.

На відміну від існуючої статичної архітектури мережі LTE EPC пропонується система, в якій службові потоки обслуговуються апаратними засобами, а у разі очікуваного перевантаження здійснюється перерозподіл потоків з урахуванням розширення мережі обслуговування за рахунок додаванням віртуальних обслуговуючих сутностей розміщених у орендованих хмарах дата центрів. Після організації гібридного обслуговуючого середовища виникає необхідність у адаптації обчислювальних ресурсів системи в процесі функціонування для забезпечення якісного обслуговування, а також необхідно врахувати особливості процесу реконфігурації та витрати, які з ним пов'язані. Досі не було комплексного рішення для задачі керування обчислювальними ресурсами гібридного телекомунікаційного середовища, не враховувалися особливості розподілу навантаження ресурсів мережевих елементів, апаратних або віртуальних.

Таким чином, науково-технічна задача реконфігурації обчислювальних ресурсів базової мережі на основі технології віртуалізації є актуальною.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалась згідно із планами науково-дослідних робіт кафедри інформаційно-телекомунікаційних мереж КПІ ім. Ігоря Сікорського у рамках держбюджетних тем №2861-п «Технологія обробки сервісів з інтеграцією інформаційних ресурсів в системах підтримки операційної діяльності підприємств зв'язку» (державний реєстраційний номер – 0115U000217), №2117-п «Технологія побудови динамічних реєстрів електронних інформаційних ресурсів та засобів їх ефективної обробки у датацентрах гетерогенної структури» (державний реєстраційний номер – 0118U003522), госпдоговірної роботи: «Розробка системної інфраструктури єдиного інформаційного середовища даних антарктичних досліджень» та ініціативної роботи: «Технологія забезпечення якості обслуговування телекомунікаційних сервісів при частковому руйнуванні фізичного обладнання» (державний реєстраційний номер – 0117U001861).

Мета і задачі дослідження. Метою дослідження є збільшення коефіцієнту використання обчислювальних ресурсів під час функціонування базової мобільної мережі (EPC) для гарантування належного рівня якості надання гібридних сервісів за рахунок динамічного перерозподілу службових потоків в мережі та реконфігурації обчислювальних ресурсів обслуговування.

Для досягнення мети дослідження було поставлено та вирішено такі **основні задачі**:

1. Проаналізувати принципи функціонування базової мережі мобільного зв'язку, особливості процесу обслуговування сервісів в ній, систематизувати методи керування обчислювальними ресурсами обслуговування в умовах гібридної інфраструктури телекомунікаційного оператора.

2. Розробити метод визначення місця розміщення та необхідної ємності віртуальних зарезервованих обчислювальних ресурсів у разі виникнення перевантаження фізичної мережі для забезпечення процесу обслуговування службових потоків у базовій мережі на належному рівні якості та підвищення коефіцієнту використання ресурсів системи, який не погіршуватиме показники обслуговування та мінімізуватиме обсяг необхідних обчислювальних ресурсів під час надання гібридних телекомунікаційних сервісів.

3. Розробити метод визначення розміру інтервалу часу сталої конфігурації ресурсів з урахуванням очікуваного навантаження в мережах з гібридною структурою, який за рахунок динамічної зміни розміру інтервалу керування ресурсами обслуговування телекомунікаційних сервісів дає змогу зменшити кількість реконфігурацій системи обслуговування та за рахунок перерозподілу обчислювальних ресурсів гарантувати задану якість обслуговування.

4. Вдосконалити метод локальної реконфігурації обчислювальних ресурсів в мережі оператора мобільного зв'язку з гібридною структурою у разі виникнення нештатної ситуації за рахунок гнучкого вибору конфігурації розміщення мережових функцій та децентралізованого контролю показників якості.

5. Провести оцінку запропонованих рішень щодо керування обчислювальними ресурсами в нормальних та нештатних умовах функціонування системи базової мобільної мережі.

Об'єктом дослідження є процес керування обчислювальними ресурсами базової мобільної мережі LTE з гібридною структурою.

Предметом дослідження є методи та моделі обробки телекомунікаційних сервісів в мережах з гібридною інфраструктурою (структура яких складається з віртуалізованих обчислювальних ресурсів та спеціалізованих ресурсів оператора зв'язку).

Методи дослідження, застосовані для вирішення поставлених завдань: методи системного аналізу, що використовуються для синтезу системи керування ресурсами, визначення структурних взаємозв'язків її компонентів та поведінки, методи математичного аналізу, теорії масового обслуговування та теорії оптимізації для визначення місць розташування мережових функцій та часу і кількості виділених їм ресурсів, математичне та імітаційне моделювання для дослідження роботи запропонованих методів реконфігурації ресурсів в гібридних мережах.

Наукова новизна отриманих результатів полягає у структурованому підході до процесу керування ресурсами мережових функцій за допомогою послідовного забезпечення контролю на етапах: моніторингу, прогнозування, контролю достатності ресурсів обслуговування телекомунікаційних сервісів. Це забезпечується такими новими науковими результатами:

1. Вперше розроблено метод визначення місця розміщення та обсягів необхідних обчислювальних ресурсів віртуальних мережових функцій, який на відміну від існуючих враховує доступні фізичні мережові ресурси та віртуалізовані обчислювальні ресурси мережі, вимоги до якості обслуговування службових потоків у EPC, що дозволяє зменшити обсяг необхідних затрат ресурсів, здійснюючи розрахунок топології та обсягу ресурсів в одній моделі.

2. Вперше запропоновано метод визначення розміру інтервалу часу сталої конфігурації обчислювальних ресурсів, який на відміну від існуючих враховує витрати на реконфігурацію системи ЕРС та забезпечення заданих показників обслуговування прогнозованого навантаження, а також подальшу оптимізацію інтервалу перерозподілу обчислювальних ресурсів залежно від динаміки навантаження, що дозволяє зменшити частоту проведення реконфігурації системи обслуговування та, за рахунок перерозподілу обчислювальних ресурсів, гарантувати задану якість обслуговування і підвищити коефіцієнт використання ресурсів.

3. Набув подальшого розвитку метод локальної реконфігурації обчислювальних ресурсів віртуальної мережі у випадку відмови або перевантаження підсистем ЕРС, який враховує динамічну зміну топології мережі обслуговування та, на відміну від існуючих, дозволяє гнучко визначати наступну конфігурацію обчислювальних ресурсів обслуговування, розміщувати мережеві функції з мінімальними витратами та децентралізовано контролювати показники якості.

Практичне значення одержаних результатів.

1. Розроблені механізми керування обчислювальними ресурсами обслуговування сервісів у гібридному середовищі були використані для обробки антарктичних даних у Національному антарктичному науковому центрі, проведені дослідження і отримані результати показали, що використання запропонованих рішень буде ефективним для вдосконалення існуючих телекомунікаційних мереж та при побудові нових перспективних гнучких добре масштабованих мереж оператора зв'язку із повним або частковим застосуванням принципу віртуалізації мережевих функцій.

2. Розроблені математичні моделі можуть бути використані при побудові блоку прийняття рішень щодо виділення обчислювальних ресурсів системи ЕРС для організації гібридного інформаційно-телекомунікаційного середовища, що гарантує якість обслуговування на заданому рівні під час обслуговування сервісів зі змінним трафіком, відповідні дослідження були проведені у лабораторіях компанії Lifecell.

3. Результати досліджень носять інтердисциплінарний характер, впроваджені у навчальний процес Інституту телекомунікаційних мереж КПІ ім. Ігоря Сікорського. Проведено апробацію та виконано практичне тестування засобів керування перерозподілом ресурсів в лабораторіях телекомунікаційних компаній Huawei та Lifecell, результати моделювання запропонованих методів показали економію ресурсів обслуговування до 15%.

Особистий внесок здобувача. Дисертаційна робота є узагальненням результатів теоретичних і експериментальних досліджень, проведених автором самостійно. Основні результати, отримані автором особисто: метод визначення місця розміщення та обсягів зарезервованих обчислювальних ресурсів віртуальних мережевих функцій, метод визначення розміру інтервалу часу сталої конфігурації обчислювальних ресурсів, метод локальної реконфігурації обчислювальних ресурсів.

Апробація результатів дисертації. Основні положення та результати

дисертаційної роботи були представлені й одержали схвалення на міжнародній конференції «The Experience of Designing and Applications of CAD Systems in Microelectronics» (м. Львів, Україна, 2015 р.), міжнародній науково-технічній конференції «Проблеми телекомунікацій» (м. Київ, Україна, 2015, 2016, 2017, 2018 рр.), міжнародній науково-практичній конференції «Інфокомунікації – сучасність та майбутнє» (м. Одеса, Україна, 2015 р.), міжнародній конференції «Black Sea Conference on Communications and Networking» (Варна, Болгарія, 2016 р.), міжнародній конференції «Radio Electronics & Info Communications» (м. Київ, Україна, 2016 р.), міжнародній науково-практичній конференції «Problems of Infocommunications. Science and Technology» (м. Харків, Україна, 2017 р.).

Публікації. За результатами досліджень опубліковано 17 наукових праць, у тому числі 1 монографія, 6 статей у наукових фахових виданнях (з них 2 у виданнях України, які включені до міжнародних наукометричних баз), 9 тез доповідей в збірниках матеріалів конференцій, 1 стаття у інших виданнях.

Структура і обсяг дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел із 122 найменувань. Загальний обсяг роботи 173 сторінки, з яких 151 сторінка основного тексту, 15 сторінок використаних джерел та 7 сторінок додатків. Робота містить 78 рисунків, 3 таблиці.

ОСНОВНИЙ ЗМІСТ ДИСЕРТАЦІЇ

У **вступі** обґрунтовано актуальність теми дисертаційної роботи, сформульовано мету і задачі дослідження, наукову новизну та практичне значення отриманих результатів. Наведено дані про впровадження результатів роботи, її апробацію, публікації та особистий внесок здобувача.

Перший розділ дисертаційної роботи присвячений аналізу стану досліджень та процесів розвитку телекомунікаційних мереж наступного покоління.

Проаналізовано ряд публікацій, що визначають місце хмарних обчислювальних ресурсів у побудові телекомунікаційних систем, які відповідають сучасним вимогам. Наявність потужних дата центрів значно розширюють можливості організації процесу надання послуг. Одним з ключових аспектів у області віртуалізації мережі є виділення фізичних ресурсів віртуальним функціям мережі. Це передбачає відображення віртуальних мереж на фізичні мережі, а також керування виділеними ресурсами протягом усього життєвого циклу віртуальної мережі. Оптимальність і гнучкість розподілу ресурсів є основними факторами для успішної віртуалізації мережі.

Виявлено, що більшість існуючих методів розв'язання задач організації апаратних та віртуалізованих ресурсів, пропонують статичну схему розподілу ресурсів, в якій при організації обчислювально-телекомунікаційного середовища перерозподіл ресурсів не відбувається протягом всього його життєвого циклу. Оскільки службовий мережевий трафік не є статичним, це може привести до неналежного використання загальних обчислювальних ресурсів. Важливо організувати моніторинг віртуальних вузлів, і на основі їх реальних потреб у ресурсах, виділяти їм їх.

Другий розділ присвячений методу визначення місця розміщення та необхідної ємності віртуальних зарезервованих обчислювальних ресурсів у разі виникнення перевантаження фізичної мережі.

Метод базується на спільному розташуванні індивідуальних віртуалізованих сервісів базової мережі на фізичній мережі. Припускаємо, що віртуальні мережеві функції мобільної базової мережі мають таку ж функціональність і інтерфейси як і мережеві компоненти архітектури 3GPP LTE Evolved Packet Core (EPC).

Кількість сервісних ланцюгів потрібно визначити заздалегідь. Частковим випадком є розгляд одного сервісного ланцюга на стільник/eNodeB. Оскільки реалістичні сценарії для мобільних мереж складають 10000 eNodeB, результуюча оптимізаційна модель буде дуже велика, і для її вирішення потрібен тривалий час на обчислення. Тому обираємо обґрунтовано великі кластери eNodeBs і припустимо, що кожен з цих кластерів eNodeB звертається до одного сервісного ланцюга базової мережі.

Розглядається випадок, коли провайдер телекомунікаційних послуг вже має існуючу топологію базових станцій. Потрібно визначити підмножину мережевих вузлів, де будуть розміщені блоки агрегації навантаження, які будуть формувати запити до одного віртуалізованого сервісу EPC. Після цього для кожного сайту базових станцій призначається вузол агрегації (Traffic Aggregation Point – TAP).

Нехай x_i – це бінарна змінна рішення, яка приймає значення 1, якщо в точці i необхідно розмістити TAP, та 0 – в іншому випадку. Крім того, визначаємо y_{ji} як бінарну змінну, яка приймає значення 1, якщо базова станція j направляє навантаження в i -у TAP, і 0 – в іншому випадку. Необхідно визначити значення x_i та y_{ji} , так щоб знайти оптимальне значення цільової функції.

Цільова функція (1) прагне мінімізувати мережеві затримки. Цільова функція (2) представляє загальну умовну вартість встановлення вузлів агрегації трафіку та вартість встановлення каналів між базовими станціями та відповідними TAP, що їх обслуговують. Цільова функція (3) має на меті залишати більше вільної пропускної здатності на кожному фізичному каналі. Максимізується залишкова пропускна здатність по всіх каналах, оскільки значно завантажені канали можуть призвести до перевантажень у мережі, тому бажано отримати рішення, де залишається більше потужностей каналів.

Ці цілі оптимізації можуть бути корисними для мережевих операторів для планування найкращої стратегії розгортання.

$$\min_{x_i, y_{ji}} (\sum_i \sum_j y_{ji} \cdot L_{ji}), \quad (1)$$

$$\min_{x_i, y_{ji}} (\sum_i x_i \cdot cost_i + \sum_i \sum_j y_{ji} \cdot cost_{lji}), \quad (2)$$

$$\max_{x_i, y_{ji}} (\sum_i \sum_j y_{ji} \cdot (c_{ji} - B_{ji})), \quad (3)$$

де L_{ji} – затримка каналу зв'язку між сайтом j і TAP i ;

$cost_i$ – вартість, яка складається з двох частин: фіксованої початкової вартості f_i , яка відповідає за фіксовані інвестиції, такі як простір та встановлення обладнання, а також додаткові витрати $costN_i$ на одиницю обробної потужності, встановленої на обчислювальному вузлі, де dr_i – обсяг задіяних обчислювальних

ресурсів обробки: $cost_i = f_i + costN_i \cdot dr_i$; $dr_i = \sum_j y_{ji} \cdot d_j$, де d_j – обсяг необхідних обчислювальних ресурсів обробки сайту j .

$costl_{ji}$ – вартість встановлення зв'язку між сайтом j і ТАР i , визначається як лінійна комбінація початкової фіксованої вартості fl_{ji} та змінної частини залежної від смуги пропускання B_{ji} , необхідної каналу, та вартості одиниці пропускну здатності $costL_{ji}$: $costl_{ji} = fl_{ji} + costL_{ji} \cdot B_{ji}$;

c_{ji} – доступна пропускну здатність.

Можливо використовувати лінійну комбінацію (4) виразів (1)-(3) з ваговими коефіцієнтами a, b, c , які можуть використовуватись не лише щоб задати більшу вагомість тої чи іншої складової, але також щоб масштабувати значення виразів з метою зведення до порівнюваних значень та мати значуще складення:

$$\min_{x_i, y_{ji}} \left(a \cdot \sum_i \sum_j y_{ji} \cdot L_{ji} + b \cdot \left(\sum_i x_i \cdot cost_i + \sum_i \sum_j y_{ji} \cdot costl_{ji} \right) - c \cdot \left(\sum_i \sum_j y_{ji} \cdot (c_{ji} - B_{ji}) \right) \right). \quad (4)$$

Обмеження:

$$\sum_i y_{ji} = 1 \quad \forall j, \quad (5)$$

$$y_{ji} \leq x_i \quad \forall j \quad \forall i, \quad (6)$$

$$\sum_i x_i \leq p, \quad (7)$$

$$\sum_j y_{ji} \cdot d_j \leq p_i \quad \forall i, \quad (8)$$

$$\sum_i y_{ji} \cdot (c_{ji} - B_{ji}) \geq 0 \quad \forall j, \quad (9)$$

$$\sum_i y_{ji} \cdot L_{ji} \leq T_j \quad \forall j. \quad (10)$$

Обмеження (5) гарантує, що кожна базова станція буде приєднана до одного ТАР. Обмеження (6) гарантує, що між сайтом базової станції j та ТАР i створюється канал, тільки якщо i було розміщено.

Обмеження (7) гарантує, що максимальна кількість ТАР не перевищує бюджет p , тоді як (8) є обмеженням потужності, що гарантує, що загальні вимоги до обробки всіх базових станцій, призначених для конкретного ТАР, не перевищують фактичні встановлені фізичні ресурси p_i . Обмеження (9) гарантує достатність каналних ресурсів для встановлення каналів, а (10) – допустимість значення затримки, тобто поріг T_j не перевищено.

Далі описано метод вирішення задачі розміщення та обсягів зарезервованих обчислювальних ресурсів віртуальних мережевих функцій.

Фізична мережа задана у вигляді графа $SN = (N, NE)$, де N є множиною фізичних вузлів і L – множиною каналів. Кожен канал $l = (n_1, n_2) \in NE$, $n_1, n_2 \in N$ має максимальну пропускну здатність $c(n_1, n_2)$ і кожен вузол $n \in N$ пов'язаний з певними ресурсами c_n^i , $i \in R$, де R – множина типів ресурсів (ядра CPU, пам'ять,

дисковий простір, мережевий інтерфейс). Множина усіх точок агрегації трафіку (ТАР), тобто кластерів eNodeB, в мережі позначається $K \subseteq N$. Для кожного вузла $n \in N$, $suit_n^{k,j}$ є бінарним параметром, який вказує, чи адміністративно можливо розгорнути на вузлі функцію типу $j \in V$, де V є множиною типів мережевих функцій, k -го сервісу, де $k \in K$.

Віртуальна базова мобільна мережа представляється множиною сервісів (один сервіс на ТАР), які вбудовуються в фізичну мережу.

Вимоги смуги пропускання каналу між двома функціями, j_1 і j_2 , $(j_1, j_2) \in E$, що відносяться до сервісу ТАР $k \in K$ позначається як $d_k^{(j_1, j_2)}$. $d_k^{j,i}$ – кількість обчислювального ресурсу типу i , що виділяється для мережевої функції j сервісу k . $s_{n,i}^{k,j}$ позначає час обробки запиту на ресурсі типу i віртуальної мережевої функції j сервісу k однією одиницею ресурсу вузла n . Вимоги до допустимого часу обробки заявки мережевою функцією j , що відноситься до сервісу k , позначаються як P_k^j . T_k – максимальна затримка для $k \in K$, $L(n_1, n_2)$ – мережева затримка для каналу $(n_1, n_2) \in NE$.

Метою оптимізації є знаходження розташування віртуалізованих сервісів базової мережі (тобто розміщення мережевих функцій та розподіл ресурсів, а також визначення шляхів передачі трафіку між ними), так щоб мінімізувати умовні витрати на зайняті ресурси каналів і вузлів у фізичній мережі, при цьому задовольняючи вимоги навантаження $\lambda^{k,j}$. Сформулюємо цільову функцію (вираз (11)) у вигляді лінійної комбінації двох вартісних виразів: зайнятого обсягу ресурсів обчислювальних вузлів, де умовна вартість одиниці ресурсу i на вузлі n позначається як $costN(i, n)$, і зайнятої пропускної здатності каналів, де $costL(n_1, n_2)$ – умовна вартість одиниці пропускної здатності фізичного каналу $(n_1, n_2) \in NE$.

Наступні вирази (11)-(20) представляють собою постановку оптимізаційної задачі нелінійного програмування. Булеві змінні $x_n^{k,j}$ вказують, чи мережева функція j пов'язана з сервісом k розташовується на фізичному вузлі n . Для $j=ТАР$, $x_n^{k,ТАР}$ – не змінні, а вхідні параметри, які вказують де ТАР k знаходиться, тобто

$$x_n^{k,ТАР} = \begin{cases} 1 & \text{якщо } k = n, \\ 0 & \text{інакше} \end{cases}.$$

Аналогічно, булеві змінні $f_{(n_1, n_2)}^{k, (j_1, j_2)}$ вказують, чи фізичний канал $(n_1, n_2) \in NE$ використовується для шляху між j_1 і j_2 для сервісу k .

$$\min_{x_n^{k,j}, f_{(n_1, n_2)}^{k, (j_1, j_2)}, d_t^{j,i}} \left(\sum_{k \in K} \sum_{j \in V} \sum_{n \in N} \sum_{i \in R} x_n^{k,j} \cdot d_k^{j,i} \cdot costN(i, n) + \sum_{(n_1, n_2) \in NE} costL(n_1, n_2) \cdot \sum_{k \in K} \sum_{(j_1, j_2) \in E} f_{(n_1, n_2)}^{k, (j_1, j_2)} \cdot d_k^{(j_1, j_2)} \right) \quad (11)$$

$$\text{З обмеженнями } \sum_{n \in N} x_n^{k,j} = 1 \quad \forall k \in K, j \in V \quad (12)$$

$$x_n^{k,j} \leq suit_n^{k,j} \quad \forall k \in K, j \in V, n \in N \quad (13)$$

$$\sum_{(w, n) \in NE} \sum_{k \in K} \sum_{(j_1, j_2) \in E} f_{(w, n)}^{k, (j_1, j_2)} \cdot d_k^{(j_1, j_2)} \leq c_n^{bdw} \quad \forall n \in N \quad (14)$$

$$\sum_{k \in K} \sum_{j \in V} x_n^{k,j} \cdot d_k^{j,i} \leq c_n^i \quad \forall n \in N, i \in \{R \setminus bdw\} \quad (15)$$

$$\sum_{k \in K} \sum_{(j_1, j_2) \in E} f_{(n_1, n_2)}^{k, (j_1, j_2)} \cdot d_k^{(j_1, j_2)} \leq c(n_1, n_2) \quad \forall (n_1, n_2) \in NE \quad (16)$$

$$\sum_{(n, w) \in NE} f_{(w, n)}^{k, (j_1, j_2)} - f_{(n, w)}^{k, (j_1, j_2)} = x_n^{k, j_1} - x_n^{k, j_2} \quad \forall k \in K, n \in N, (j_1, j_2) \in E \quad (17)$$

$$x_n^{k, j}, f_{(n_1, n_2)}^{k, (j_1, j_2)} \in \{0, 1\} \quad \forall k \in K, j \in V, n \in N, (j_1, j_2) \in E, (n_1, n_2) \in NE \quad (18)$$

$$\sum_{(j_1, j_2) \in E} \sum_{(n_1, n_2) \in L} f_{(n_1, n_2)}^{k, (j_1, j_2)} \cdot L(n_1, n_2) \leq T_k \quad \forall k \in K \quad (19)$$

$$\sum_{n \in N} x_n^{k, j} \sum_{i \in R} \left(\frac{1}{\frac{d_k^{j,i}}{s_{n,i}^{k,j}} - \lambda^{k,j}} \right) \leq P_k^j \quad \forall t \in T, j \in V \quad (20)$$

Вираз (12) гарантує, що для кожної ТАР/сервісу розміщується тільки одна мережева функція кожного типу. Вираз (13) гарантує, що розміщення ресурсів здійснюється на фізичних вузлах, які мають адміністративну можливість для розташування відповідних мережевих функцій. Вирази (14), (15) і (16) являють собою обмеження на доступні ресурси фізичних вузлів і каналів. Вираз (17) представляє собою обмеження щодо збереження потоку для всіх шляхів у фізичній мережі. Вираз (18) гарантує, що змінні у задачі розміщення функцій мережі та відображення шляху є булевими.

Щоб обмежити затримки на каналах, обмеження на затримку, показане в виразі (19), також додається. А для того, щоб урахувати у моделі необхідну продуктивність віртуальної мережевої функції, необхідні обмеження на значення часу обробки заявки, що визначені у виразі (20).

Передбачається вирішення задачі (11)-(20) в офлайн режимі на початковому етапі. Згідно з рішенням, кожній мережевій функції резервується певна кількість ресурсів віртуальної мережевої функції, на основі оцінки її найбільшої потреби в ресурсах; миттєві потреби різних мережевих функцій динамічно задовольняються шляхом активації необхідної конфігурації віртуальних машин під час виконання таким чином, щоб задовольнити гарантії передбачені для кожної мережевої функції.

Кількісний та якісний аналіз (рис. 1) запропонованого методу показав скорочення умовних витрат пов'язаних з зарезервованими ресурсами до 15%, що сприяє підвищенню ефективності обслуговування навантаження, економії обчислювальних ресурсів.

В третьому розділі запропоновано метод визначення розміру інтервалу часу сталої конфігурації обчислювальних ресурсів.

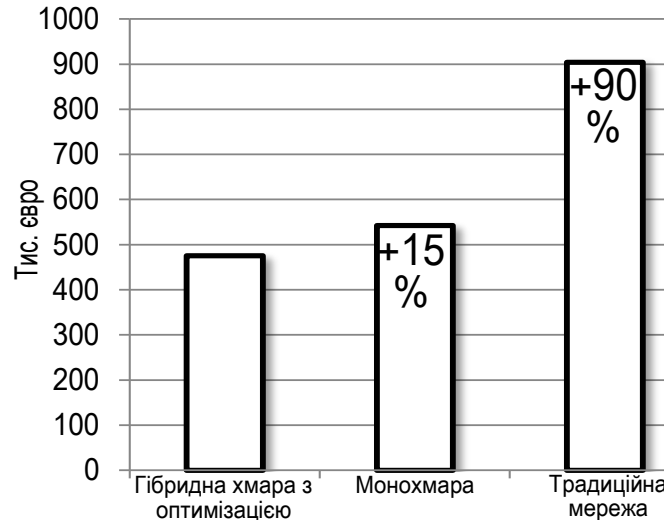


Рисунок 1 – Витрати ресурсів системи із застосуванням методу розміщення віртуальних мережеских функцій та без нього

Рішення про те, коли надавати ресурси, залежить від динаміки навантажень. Телекомунікаційні навантаження зазнають довгострокових змін, таких як вплив години дня або сезонні ефекти, а також короткострокових коливань таких як неочікувані натовпи. У той час як довгострокові коливання можуть бути передбачені заздалегідь, спостерігаючи за змінами в минулому, короткострокові коливання менш передбачувані, а в деяких випадках, не передбачувані. Запропонована методика використовує два різних методи для роботи в умовах змін, які спостерігаються в різних часових масштабах. Використовується прогностичне керування ресурсами для оцінки навантаження і відповідного керування, а також реактивне керування ресурсами для виправлення помилок у довгострокових прогнозах або для реагування на непередбачені перенавантаження.

Запропоновано механізм, суть якого полягає у динамічній зміні тривалості сталої конфігурації ресурсів віртуальної мережескої функції залежно від різниці між максимальним значенням навантаження на базовому інтервалі та мінімальним. Формула (21) описує принцип:

$$Int(t) = \max \left(Int_{base} \cdot \left(1 - K \cdot \frac{\max_{\tau \in (t-I(t-1);t)} \lambda_{basepred}(\tau) - \min_{\tau \in (t-I(t-1);t)} \lambda_{basepred}(\tau)}{\max \lambda_{basepred}} \right); Int_{min base} \right), \quad (21)$$

де Int – інтервал, протягом якого будуть виділятися відповідні визначені ресурси, Int_{base} – базове значення інтервалу, яке розраховується відповідно до описаного далі підходу до дискретизації навантаження, K – коефіцієнт зміни тривалості сталої конфігурації, яка визначається оператором мережі відповідно до експерименту, $\lambda_{basepred}(t)$ – середньостатистична передбачена інтенсивність надходження заявок в період t , $Int_{min base}$ – мінімальна допустима величина базового інтервалу.

Для цього потрібно визначити базовий інтервал. Мета полягає в тому, щоб представити добовий шаблон в навантаженні, дискретизуючи його запити у

послідовні, непересічні часові інтервали з єдиним репрезентативним значенням в кожному інтервалі. Дискретизація навантаження: маючи часовий ряд X на області $[v, \tau]$, часовий ряд Y на тій же самій області є дискретизацією X , якщо $[v, \tau]$ може бути розділене на m послідовних непересічних часових інтервалів, $\{[v, \tau_1], [\tau_1, \tau_2], \dots, [\tau_{m-1}, \tau]\}$, так що $X(j)=r_i$, для всіх j у i -му інтервалі, $[\tau_{i-1}, \tau_i]$.

Пропонується рішення для дискретизації часових рядів (вираз 22):

$$\sum_{i=1}^m [\sum_{\tau=\tau_{i-1}}^{\tau_i} u(r_i - X(\tau))] + f(m) \rightarrow \min . \quad (22)$$

Вираз (22) є цільовою функцією, яку потрібно мінімізувати, де X являє собою часовий ряд і $f(m)$ є функцією вартості кількості змін або інтервалів, m . Метою виразу (22) є одночасна мінімізація помилки репрезентації навантаження і кількості змін. Базовий інтервал обчислюємо як $Int_{base} = \frac{\tau_m}{m}$. Для визначення оптимальної величини інтервалу задано різні значення кількості інтервалів, після обчислення значення виразу (22) і обирається найкраще, тобто мінімальне, при цьому на кожному інтервалі:

$$r_i = \max_{\tau \in (\tau_{i-1}, \tau_i)} X(\tau). \quad (23)$$

При цьому пропонується здійснювати постійний моніторинг значень інтенсивності навантаження та використовувати прогнозовані значення, якщо інтенсивність навантаження не перевищує порогової, в іншому випадку – оцінювати поточні тренди та здійснювати масштабування ресурсів на основі нового прогнозу.

Прогнозування навантаження на наступний інтервал часу здійснюється шляхом урахування довгострокової статистики та корегуванням його за моделлю експоненційного згладжування, де помилки більш нових минулих періодів мають більший ваговий коефіцієнт:

$$\lambda_{pred}(t) = \lambda_{basepred}(t) + \alpha \sum_{j=t-h}^{t-1} (1 - \alpha)^{t-1-j} \cdot (\lambda_{obs}(j) - \lambda_{basepred}(j))^+, \quad (24)$$

де α (константа згладжування) – коефіцієнт, що характеризує швидкість зменшення ваг, та приймає значення від 0 до 1, чим значення цього параметра ближче до одиниці, тим більше при прогнозі враховується вплив останніх рівнів ряду. Параметри моделі встановлюються адміністратором мережі відповідно до експерименту. $\lambda_{obs}(t)$ – інтенсивність надходження навантаження протягом інтервалу t , h – інтервал минулих спостережень, який розглядається алгоритмом, а x^+ позначає $\max(0, x)$.

Приклади представлення значень часового ряду, тобто навантаження, які ілюструють точність репрезентації залежно від обраного інтервалу сталої конфігурації, представлені на рис. 2, де помилка репрезентації для випадку інтервалів у 10 хвилин складає 7%, а для випадку 60 хвилин – вже 19%.

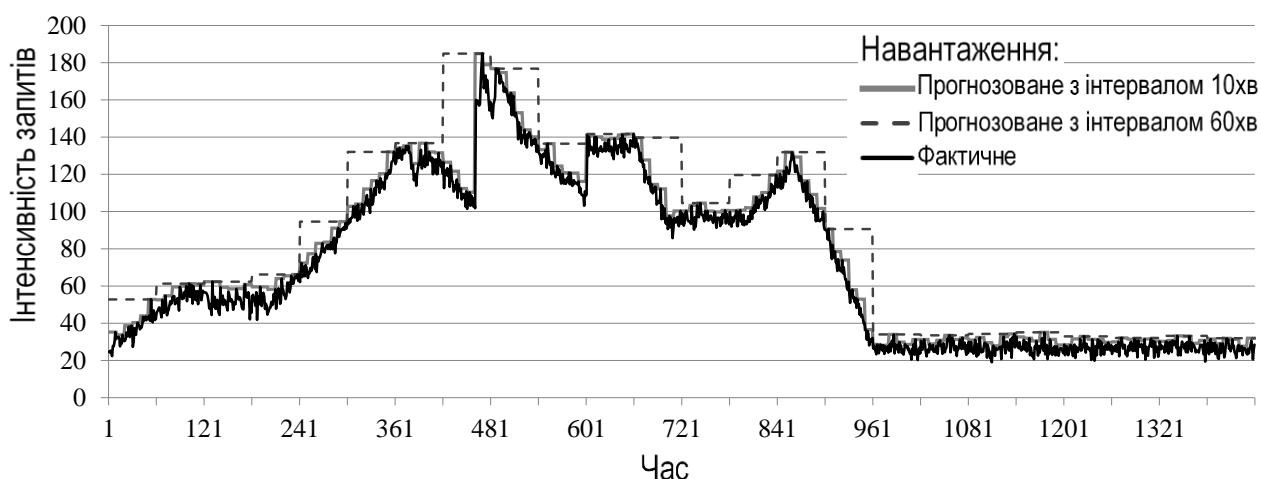


Рисунок 2 – Представлення значень навантаження в залежності від різної величини інтервалу сталої конфігурації

Результати моделювання методу визначення розміру інтервалу часу сталої конфігурації ресурсів (рис. 3) показали, що різниця між репрезентаційним значенням та реальним може складати 9%. Якщо не застосовувати систему динамічного регулювання величини інтервалу сталої конфігурації, то відхилення складатиме 18%, тобто на 9% більше і ресурсів буде витрачатись, відповідно, більше.

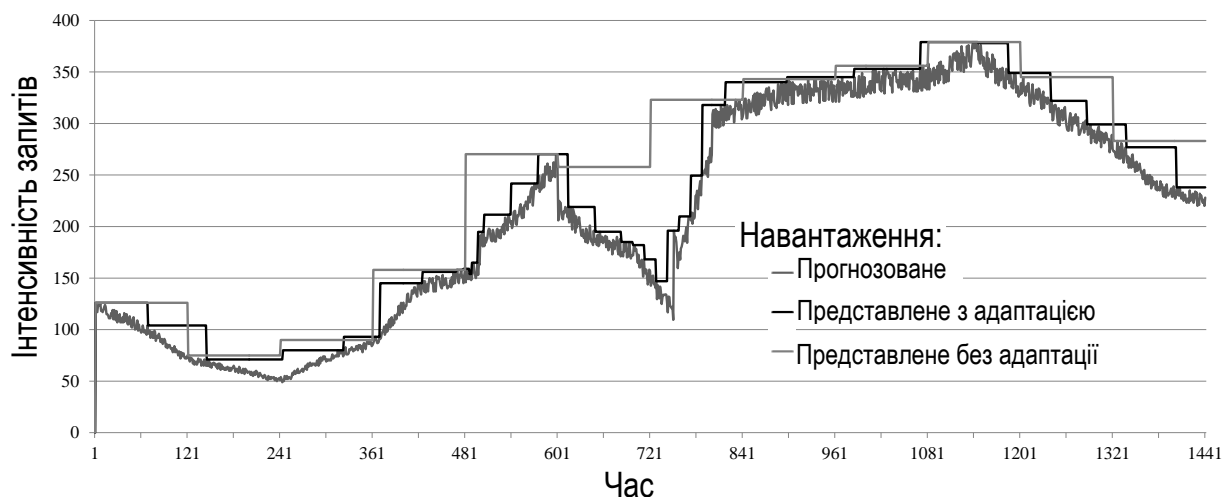


Рисунок 3 – Результати моделювання системи з динамічною зміною величини інтервалу часу сталої конфігурації та системи без неї

Для оцінки запропонованого підходу середня кількість вільних ресурсів в день визначалась як різниця між фіксовано виділеними ресурсами, тобто коли протягом доби завжди виділено 100% ресурсів, та динамічно виділеними ресурсами з використанням NFV. Відповідно до результатів імітаційного моделювання об'єм ресурсів, що виділяються динамічно, в середньому на 42% менше, ніж у разі використання традиційного підходу розподілу. На рис. 4 зображено результат роботи динамічного розподілу ресурсів у віртуалізованому ЕРС мобільної мережі у графічному вигляді. Сіра штрихова лінія ілюструє випадок фіксовано виділених ресурсів на найгірший випадок. Чорна крива показує обсяг ресурсів, розподілених відповідно до запропонованого методу динамічно.

У четвертому розділі розглядається задача локальної реконфігурації обчислювальних ресурсів мережі при відмові або перевантаженні.

Можуть виникати випадки, коли наявних на вузлах ресурсів буде недостатньо або якщо вузол відмовить. Потенційними джерелами відмов можуть бути: відмови фізичних вузлів, відмови серверів, інтенсивність відмов серверів може бути вища за телекомунікаційне апаратне забезпечення, також необхідно врахувати технічне обслуговування вузлів провайдером інфраструктури, що потребуватиме міграції вузлів.

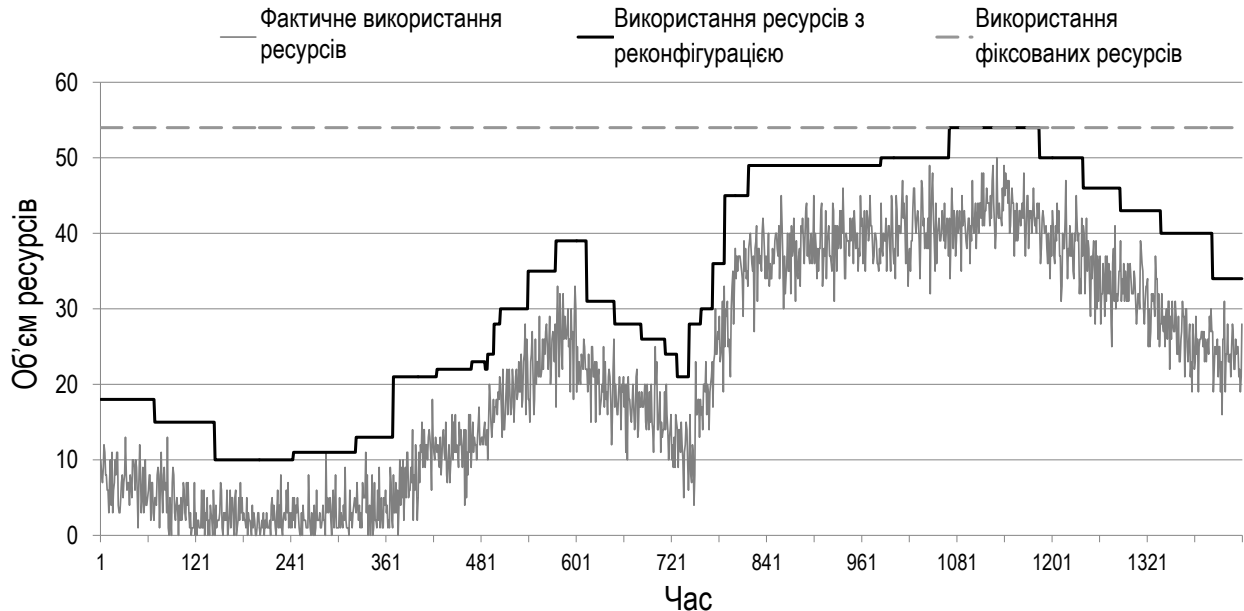


Рисунок 4 – Результати моделювання системи зі змінною конфігурацією ресурсів та системи з фіксовано виділеними ресурсами

Для такого випадку застосовуються методи реконфігурації, які вирішують завдання: знайти місця міграції мережеских функцій з постраждалих вузлів, мінімізуючи вартість відновлення вузла після відмови, а також підтримуючи високий рівень фізичної працездатності мережі. Запропоновані вдосконалені методи відновлення відрізняються від існуючих тим, що враховується вартість ресурсів на вузлах і кінцева якість обслуговування, а також випадок відмови вузла через надмірне навантаження, що надходить на нього, крім того, невирішеною залишалося завдання розташування вузлів керування – менеджерів, які координують переміщення віртуальної мережескої функції.

MN являє собою підмножину вузлів керування (далі – менеджери), де $MN \subseteq N$, які відповідають за функціонування запропонованого механізму відновлення після відмови. Кожен керуючий вузол пов'язаний з одним або декількома вузлами фізичної мережі і виконує кроки, необхідні для відновлення після відмови мережі. Припускаємо, що менеджери можуть розміщуватись в вузлах N . При заданій кількості менеджерів A існує скінченна множина з $\binom{|N|}{A}$ можливих розташувань, відповідно, задача розміщення менеджерів є задачею багатокритеріальної комбінаторної оптимізації. Метою задачі є визначення місця розташування кожного менеджера при заданій їх кількості A , так що мінімізується функція загальних витрат $U_A(\{p_n: n \in N\})$, де p_n – булева змінна, яка рівна одиниці, якщо менеджер розміщується в точці n . Задача оптимізації буде мати вигляд:

$$\min_{\{p_n: n \in N\}} U_A \quad \text{при обмеженні } \sum_n p_n = A \quad (25)$$

Основною метою оптимального розміщення менеджерів є мінімізація затримок між вузлами і менеджерами в мережі. Проте, розглядати тільки затримки недостатньо. Розміщення менеджерів повинно також враховувати певні обмеження стійкості. На рис. 5 показані різні питання, які необхідно враховувати при оцінці стійкості розміщення. Нижче коротко пояснимо ці питання і що необхідно, щоб бути стійким по відношенню до них. На рис. 5 показано нормалізовані затримки між вузлами та інтенсивність навантаження на вузлах.

Припустимо, що вузли призначаються до їх найближчого менеджера, використовуючи в якості метрики затримки, тобто найкоротший шлях $dl_{g1,g2}$ між вузлом $g1$ і менеджером $g2$. Кількість вузлів на менеджера може бути незбалансованою – чим більше вузлів менеджер повинен контролювати, тим вище навантаження на цього менеджера. Якщо кількість запитів вузла до менеджера в мережі збільшується, аналогічно поводить себе і ймовірність додаткових затримок через черги в системі керування. Для того, щоб бути стійким від перевантаження менеджера, призначення вузлів різним менеджерам повинно бути добре збалансованим.

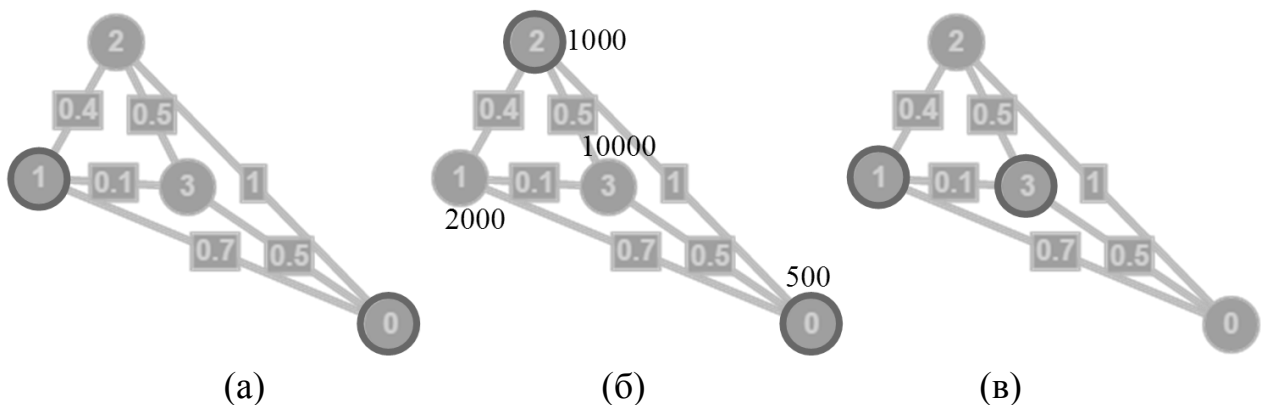


Рисунок 5 – Розміщення по різним критеріям: (а) мінімальної затримки до менеджера; (б) мінімального дисбалансу навантаження на менеджерів; (в) мінімальної затримки між менеджерами

Однак стає очевидно, що одного менеджера не досить, щоб досягти стійкості в мережі. Проте, коли кілька менеджерів розміщуються в мережі, логіка керування мережі розподіляється по декількох менеджерах і ці менеджери повинні синхронізуватися, щоб підтримувати несуперечний глобальний стан. Залежно від частоти синхронізації між менеджерами, затримка між окремими менеджерами грає важливу роль.

На основі матриці dl , що містить відстані найкоротших шляхів між усіма вузлами, максимальний час затримки передачі між вузлом і менеджером для певного розміщення менеджерів може бути визначений як:

$$U_A^{latency}(p) = \max(ddc_n),$$

де ddc_n – максимальна затримка передачі від вузла мережі до менеджера в точці n , яке розраховується наступним чином:

$$ddc_n = \max_{g \in N} latency_g \cdot \pi_{g,n},$$

де $latency_g$ – затримка між менеджером та вузлом g , $latency_g = \min_{\{n: n \in N \cap p_n = 1\}} dl_{g,n}$;

$\pi_{g,n}$ – булева змінна, яка рівна одиниці, якщо вузол g обслуговується менеджером розміщеним в точці n .

Розглядаємо не середнє, але максимальне значення затримки, так як середнє приховує значення найгіршого випадку, які є важливими, коли стійкість повинна бути поліпшена.

Залежно від ситуації, може бути бажано мати приблизно рівне навантаження на всіх менеджерів, так що жоден менеджер не перевантажується, а у інших мало роботи. Далі розглядаємо збалансований розподіл вузлів між менеджерами. Як формальну метрику вводимо баланс розміщення або, вірніше, дисбаланс, $U_A^{imbalance}$, тобто відхилення від повністю збалансованого розподілу, як різниця між навантаженням на найбільш завантаженому менеджері і найменш завантаженому менеджері.

$U_A^{imbalance}$ визначається в такий спосіб:

$$U_A^{imbalance}(p) = \max(ldc_n) - \min(ldc_n), \text{ де } ldc_n > 0,$$

де ldc_n – навантаження на менеджера в точці n , яке розраховується наступним чином:

$$ldc_n = \sum_{g \in N} load_g \cdot \pi_{g,n},$$

де $load_g$ – коефіцієнт навантаження на вузол g .

Як останній аспект стійкого розміщення менеджерів, розглянемо як затримка між менеджерами може враховуватися при виборі розміщення менеджерів. Формально, затримка між менеджерами $U_A^{interlatency}$ визначається як найбільша затримка між будь-якими двома менеджерами при заданому розміщенні:

$$U_A^{interlatency}(p) = \max_{\{n,g:n,g \in N \cap p_n = 1, p_g = 1\}} dl_{g,n}.$$

Загалом, розміщення з урахуванням затримки між менеджерами мають тенденцію розміщувати всіх менеджерів набагато ближче один до одного. Це збільшує максимальну затримку від вузлів до менеджерів.

Таким чином, цільова функція оптимізації:

$$U_A = w_l latency \times U_A^{latency}(p) + w_i imbalance \times U_A^{imbalance}(p) + w_{il} interlatency \times U_A^{interlatency}(p),$$

де w_i – множина вагових коефіцієнтів.

У задачі міграції, фізична мережа задана у вигляді графа $SN=(N,NE)$, де N є множиною фізичних вузлів і NE – множиною каналів. Кожен канал $(n_1, n_2) \in NE$, $n_1, n_2 \in N$ має максимальну пропускну здатність $c(n_1, n_2)$ і мережеву затримку

$L(n_1, n_2)$, а кожен вузол $n \in N$ пов'язаний з певними ресурсами c_n^i , $i \in R$, де R – множина типів ресурсів. Мережа зв'язку представлена множиною сервісів (або запитів віртуальної мережі) K , які вбудовуються в фізичну мережу. Запит віртуальної мережі k , $k \in K$, можна представити як зважений граф $G_k = (V_k, E_k)$, де V_k є множиною віртуальних вузлів, який містить h_k елементів і позначається як $V_k = (v_{k,1}, v_{k,2}, \dots, v_{k,h_k})$, де $v_{k,j}$ означає j -у мережеву функцію у ланцюзі функцій k , E_k є множиною віртуальних каналів $e_k(v_{k,j}, v_{k,g}) \in E_k$. Вимоги смуги пропускання каналу між двома функціями, $j1$ і $j2$, що відносяться до сервісу $k \in K$ позначається як $d_k^{(j1,j2)}$, $d_k^{j,i}$ – кількість ресурсу типу i , що виділяється для мережевої функції j сервісу k . Булеві змінні $x_n^{k,j}$ вказують, чи мережева функція j , пов'язана з $k \in K$, розташовується на фізичному вузлі n , змінні $f_{(n1,n2)}^{k,(j1,j2)}$ визначають, чи фізичний канал $(n1,n2)$ використовується у шляху між $j1$ та $j2$ для запиту k . L_k – максимальна затримка для запиту k . $costN(i,n)$ – вартість зайнятої одиниці ресурсу i на фізичному вузлі n , і $costL(n_1,n_2)$ – вартість зайнятої одиниці пропускної здатності на фізичному каналі $(n_1,n_2) \in NE$. $suit_n^{k,j}$ означає, що функція j з запиту k може бути розміщена на вузлі n .

Процес переміщення вузлів віртуальної мережі, розміщених на вузлі, який відмовив, $v_{k,j}^{fail}$, запускається, коли система відправляє запит на відновлення відповідному вузлу-менеджеру, та протікає в такий спосіб: менеджер направляє запит на відновлення до всіх вузлів фізичної мережі, на яких розміщаються віртуальні вузли, суміжні з ураженими віртуальними вузлами. Кожен з цих вузлів будує дерево найкоротших шляхів (Shortest Path Tree – SPT) до всіх вузлів фізичної мережі на відстані не більше l (поріг встановлюється провайдером послуг) від вузла, де коренем SPT виступає сам цей вузол. Менеджер використовує ці шляхи, щоб вибрати вузол з оптимальною відстанню до всіх вузлів фізичної мережі, де розташовані вузли віртуальної мережі прилеглі до несправного вузла. Цей вузол в кінцевому рахунку стає оптимальним кандидатом для розміщення ураженого віртуального вузла. Крім того, ємність кінцевих вузлів шляхів з SPT повинна бути не менше ємності віртуального вузла, розміщеного на несправному вузлі. Обираємо вузол з мінімальною вартістю шляху до всіх кореневих вузлів у деревах SPT та мінімальною вартістю обчислень. Алгоритм 1 (рис. 6) містить опис псевдокоду алгоритму відновлення вузла після відмови і виконується для всіх $\{v_{k,j} : x_n^{k,j} = 1 \text{ \& } n = \text{failed}\}$.

У мережі існує ймовірність відмови вузла через перевантаження. Для виконання відновлення при відмові вузла в перевантаженій мережі, виконується процедура реконфігурації для міграції віртуальних вузлів, розміщених на перевантаженому фізичному вузлі.

Процес відновлення починається з сортування всіх віртуальних вузлів, розміщених на перевантаженому фізичному вузлі. Критерієм, що використовуються для сортування цих вузлів віртуальної мережі, є ємність віртуальних вузлів. Потім виконується процедура відновлення на першому відсортованому вузлі віртуальної мережі, що має ємність рівну перевантаженню, для переміщення на новий вузол фізичної мережі.

Коли навантаження або ресурси змінюються, деякі віртуальні мережеві функції (VNF) можливо доведеться перемістити. Існує ймовірність того, що знайти новий вузол-кандидат для вузла віртуальної мережі, розміщеного на вузлі з відмовою, не вийде. В такому випадку виконується процедура реконфігурації для міграції одного або декількох віртуальних вузлів для переміщення розміщених вузлів віртуальної мережі. Розглянемо задачу міграції як задачу оптимізації, яка спрямована на мінімізацію загальних умовних витрат міграції при обмеженнях допустимої затримки і обчислювальних ресурсів.

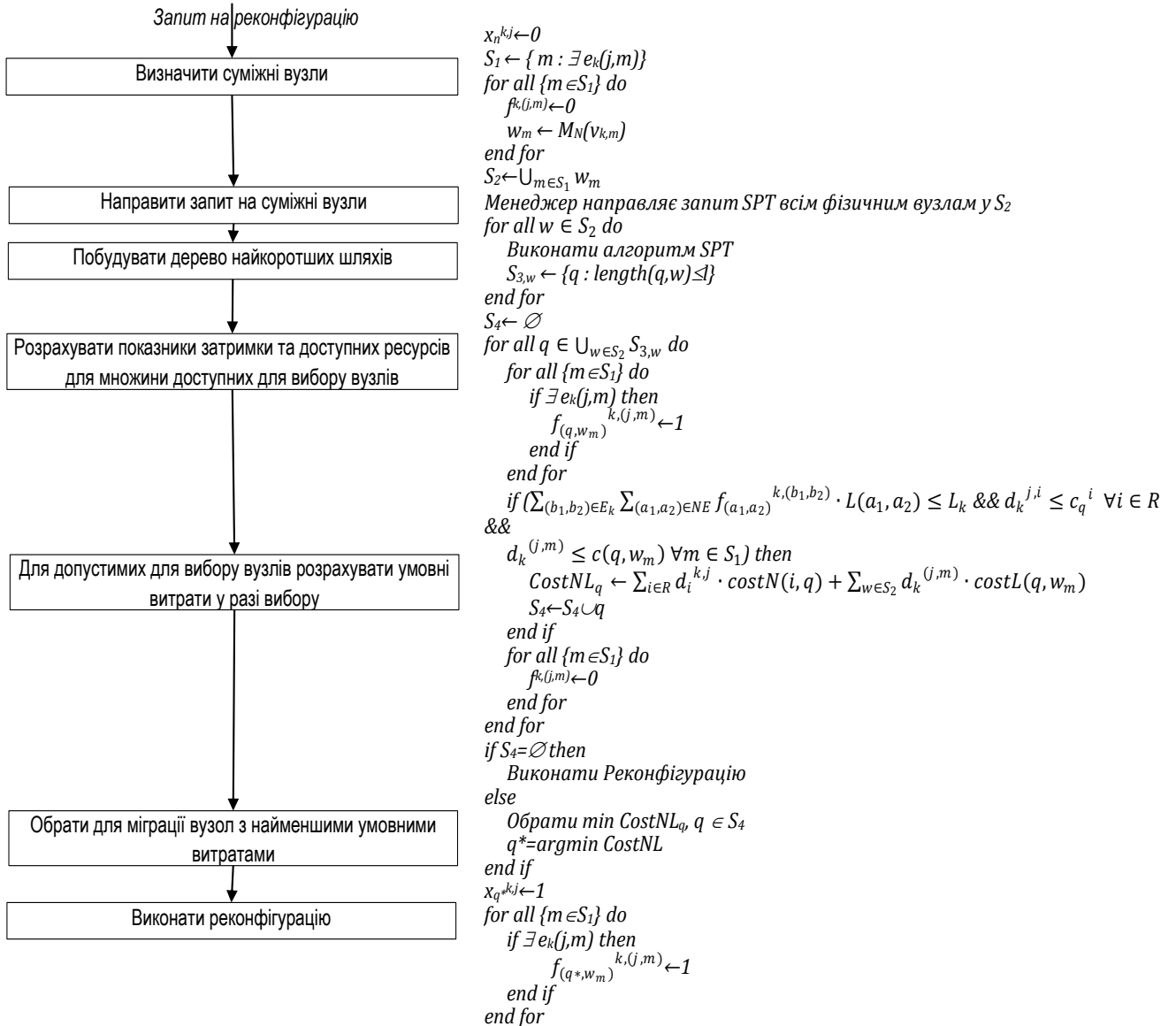


Рисунок 6 – Відновлення вузла з відмовою

Метою оптимізації є знаходження розташування віртуальних мережевих функцій (тобто розміщення мережевих функцій та розподіл ресурсів, а також визначення шляхів передачі трафіку між ними), так щоб мінімізувати умовні витрати на зайняті ресурси каналів і вузлів у фізичній мережі, при цьому задовільняючи вимоги трафіку. Сформулюємо цільову функцію (26) у вигляді лінійної комбінації (з ваговими коефіцієнтами a, b, c, e) вартісних виразів.

Визначимо бінарну змінну $x_n^{k,j} \in \{0,1\}$, для позначення того, що VNF j пов'язана з ланцюгом сервісу k розміщується на вузлі n після міграції. Індикатор

$x_n^{k,j}=0$ означає, що VNF j не розміщується на вузлі n після міграції; в іншому випадку j розміщується на вузлі n після міграції.

Введемо бінарну змінну $y_n^{k,j}$ для індикації стану мережі перед міграцією. Вона схожа зі змінною $x_n^{k,j}$, $y_n^{k,j}=0$ означає, що VNF j сервісу k не перебуває на вузлі n до міграції; в іншому випадку, j розташована на вузлі n до міграції.

Таким чином, можемо використовувати індикатор $I^{k,j}$, щоб вказати чи VNF j сервісу k було переміщено в поточному процесі міграції.

$$I^{k,j} = \sum_{n \in N} x_n^{k,j} \cdot y_n^{k,j}$$

$I^{k,j} = 0$ вказує, що VNF було переміщено в поточному процесі міграції, і $I^{k,j} = 1$ вказує, що VNF не було переміщено.

x_n позначає чи n -ий фізичний сервер працює або ні після міграції.

$$x_n = \begin{cases} 1 & (\text{сервер } n \text{ працює}) \\ 0 & (\text{в іншому випадку}) \end{cases}$$

y_n позначає чи n -ий фізичний сервер працює або ні перед міграцією.

$$y_n = \begin{cases} 1 & (\text{сервер } n \text{ працює}) \\ 0 & (\text{в іншому випадку}) \end{cases}$$

Для того щоб розглянути ресурси, які споживаються при міграції та запуску серверів, вводимо такі вирази:

– B_n позначає необхідні витрати b_n для запуску n -ого серверу:

$$B_n = b_n x_n (x_n - y_n);$$

– $L_i^{k,j}(n \rightarrow n')$ позначає використання ресурсу i для міграції VNF j з ланцюгу сервісу k з серверу n на сервер n' :

$$L_i^{k,j}(n \rightarrow n') = l_i(d^{k,j}) + l'_i(d^{k,j}),$$

де $l_i(x)$ – функція використання ресурсу i для міграції з серверу, $l'_i(x)$ – функція використання ресурсу i для міграції на сервер.

Цільова функція буде визначатися як:

$$\begin{aligned} MCost = & a \cdot \sum_{n \in N} (B_n + x_n \cdot cost(n)) + b \cdot \sum_{n \in N} \sum_{k \in K} \sum_{j \in V} \sum_{i \in R} x_n^{k,j} \cdot d_i^{k,j} \cdot \\ & costN(i, n) + c \cdot \sum_{(n_1, n_2) \in NE} costL(n_1, n_2) \cdot \sum_{k \in K} \sum_{(j_1, j_2) \in E} f_{(n_1, n_2)}^{k, (j_1, j_2)} \cdot d_k^{(j_1, j_2)} + \\ & e \cdot \sum_{n \in N} \sum_{n' \in N} L_i^{k,j}(n \rightarrow n') x_{n'} (x_{n'} - y_n) \end{aligned} \quad (26)$$

Використовуючи наведені вище міркування, формулюємо задачу наступним чином.

Цільова функція:

$$\text{Min } MCost$$

З обмеженнями:

$$\sum_{n \in N} x_n^{k,j} = 1 \quad \forall k \in K, j \in V, \quad (27)$$

$$x_n^{k,j} \leq \text{suit}_n^{k,j} \quad \forall k \in K, j \in V, n \in N, \quad (28)$$

$$\sum_{k \in K} \sum_{j \in V} x_n^{k,j} \cdot d_i^{k,j} + y_n^{k,j} \cdot (1 - I^{k,j}) \cdot l_i(d^{k,j}) + x_n^{k,j} \cdot (1 - I^{k,j}) \cdot l'_i(d^{k,j}) \leq c_n^i \quad \forall n \in N, i \in R, \quad (29)$$

$$\sum_{t \in K} \sum_{(j_1, j_2) \in E} f_{(n_1, n_2)}^{k, (j_1, j_2)} \cdot d_k^{(j_1, j_2)} \leq c(n_1, n_2) \quad \forall (n_1, n_2) \in NE, \quad (30)$$

$$\sum_{(n, w) \in L} f_{(w, n)}^{k, (j_1, j_2)} - f_{(n, w)}^{k, (j_1, j_2)} = x_n^{k, j_1} - x_n^{k, j_2} \quad \forall k \in K, n \in N, (j_1, j_2) \in E, \quad (31)$$

$$x_n^{k,j}, f_{(n_1, n_2)}^{k, (j_1, j_2)} \in \{0, 1\} \quad \forall k \in K, j \in V, n \in N, (j_1, j_2) \in E, (n_1, n_2) \in NE, \quad (32)$$

$$\sum_{(j_1, j_2) \in E} \sum_{(n_1, n_2) \in NE} f_{(n_1, n_2)}^{k, (j_1, j_2)} \cdot L(n_1, n_2) \leq L_k \quad \forall k \in K, \quad (33)$$

$$\sum_{n \in N} x_n^{k,j} \sum_{i \in R} \left(\frac{1}{\frac{d_k^{j,i}}{s_{n,i}^{k,j}} - \lambda^{k,j}} \right) \leq P_k^j \quad \forall k \in K, j \in V. \quad (34)$$

Отже, цільова функція (26) представляє собою лінійну комбінацію чотирьох вартісних виразів та направлена на мінімізацію: вартості запуску та використання серверу, використання ресурсів серверу, використання каналів зв'язку та використання ресурсів для міграції. Обмеження (27) гарантує однократність розміщення мережових функцій, а (28) – адміністративну можливість розміщення на вузлі. Вирази (29) та (30) являють собою обмеження на ресурси фізичних вузлів і каналів, тобто забезпечують той факт, що кількість задіяних на вузлі ресурсів не перевищує кількості доступних. Вираз (31) представляє собою обмеження щодо збереження потоку для всіх шляхів у фізичній мережі, тобто що вхідний потік на вузлі дорівнює вихідному потоку. Вираз (32) гарантує, що змінні у задачі є булевими. Вирази (33) та (34) представляють собою обмеження на час передачі телекомунікаційними каналами та час обробки вузлами обслуговування відповідно, і забезпечують дотримання заданих часових вимог до обслуговування сервісу.

За результатами моделювання (рис. 7) запропонований метод локальної реконфігурації показав до 27% менші умовні витрати у порівнянні зі стратегією направленою на мінімізацію затримки, при цьому затримка знаходилась у допустимих межах але була на 20% більшою.

Таким чином, перед початком роботи необхідно мати статистичні дані інтенсивності надходження заявок на мережову функцію та ймовірнісні характеристики обслуговування заявки. Згідно з методом розміщення визначається прив'язка кожної мережової функції традиційної мережі до дата центру та кількості ресурсів, що потрібно зарезервувати для відповідної віртуалізованої мережової функції. Далі необхідно розбити період роботи мережової функції на інтервали протягом яких її конфігурація буде незмінною і відповідно буде активізовано певний обсяг ресурсів згідно з методом визначення

розміру інтервалу часу сталої конфігурації ресурсів з урахуванням очікуваного навантаження. Під час функціонування мережі мобільного зв'язку може виникнути ситуація, коли фізичний вузол не в змозі надалі обслуговувати вхідне навантаження через недостатність ресурсів для цього або через вихід його з ладу, а такому випадку запускається розподілений метод локальної реконфігурації ресурсів, який перерозподіляє віртуальні вузли у мережі, що постраждали від відмови або перевантаження на фізичному вузлі.

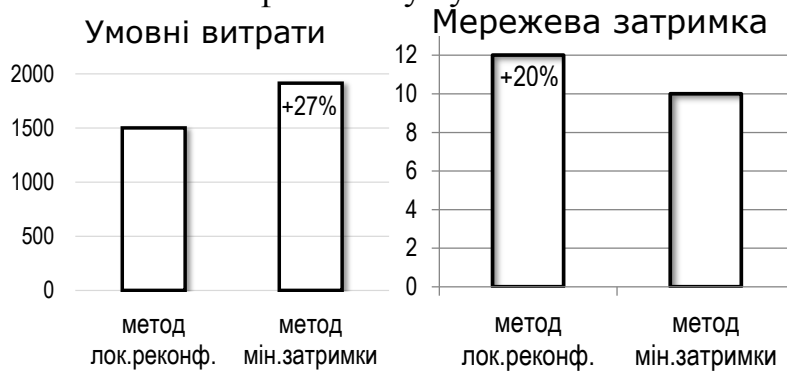


Рисунок 7 – Результати моделювання системи зі змінною конфігурацією ресурсів та системи з фіксовано виділеними ресурсами

Загальна система керування ресурсами зображена на рис. 8.

Система моніторингу відслідковує трафік і підраховує кількість запитів. В системі моніторингу встановлюється порогове значення для числа запитів і до координатора посиляється повідомлення, якщо система моніторингу виявляє перевантаження. Коли координатор приймає повідомлення про перевантаження, з системи моніторингу, блок розподілу ресурсів обчислює необхідну кількість ресурсів для обробки заявок належним чином і динамічно розподіляє розрахунковий об'єм. Потім координатор перенаправляє заявки і перевантаження усувається.

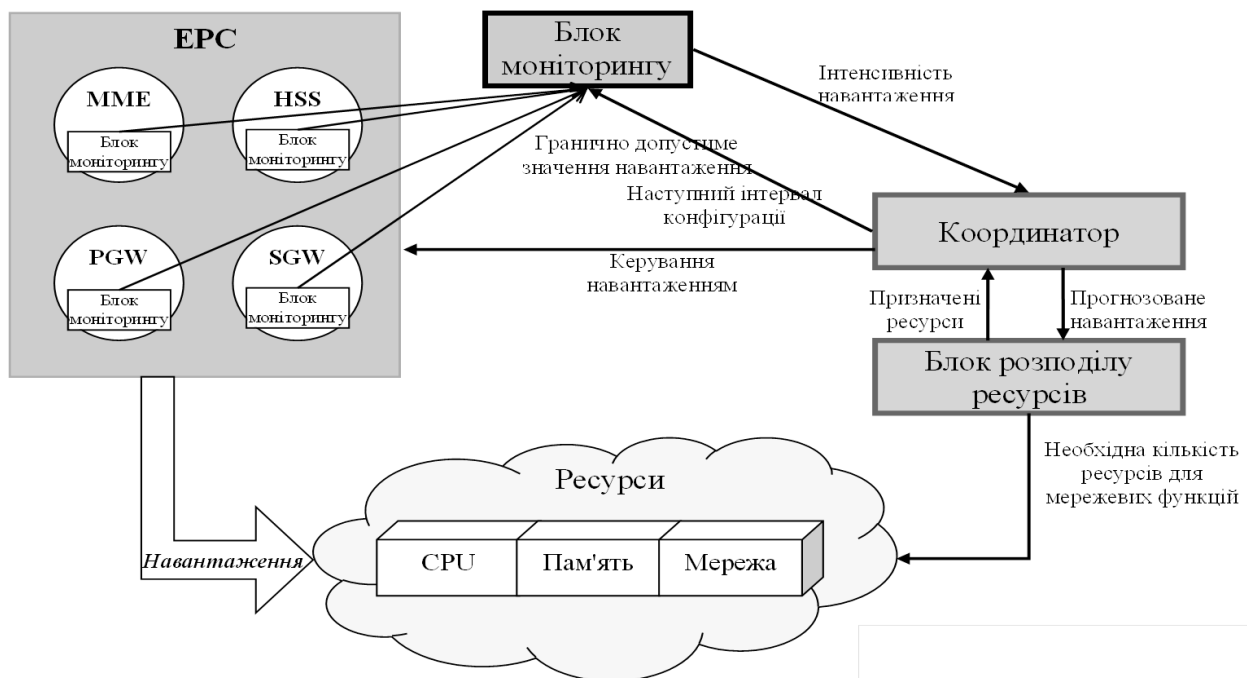


Рисунок 8 – Модифікована система керування ресурсами

Координатор запускається періодично. Для прогнозу базового навантаження, можемо взяти середнє значення історичного денного навантаження. Координатор направляє вхідне навантаження на дата центр, виділений для обслуговування надлишкового службового навантаження, а також обмінюється даними з блоком розподілу ресурсів для надання інформації про передбачене вхідне навантаження.

Блок розподілу ресурсів відповідає за розподіл належної кількості ресурсів, необхідної для обробки навантаження з заданими показниками якості. Протягом безпосереднього функціонування системи цей блок запускається коли реальне навантаження перевищує базове прогнозоване значення навантаження, для того щоб надати додаткові ресурси для надлишкового навантаження. Оскільки блок розподілу ресурсів та координатор не запускаються коли реальне навантаження нижче ніж прогнозоване, процедура реконфігурації ресурсів створює мінімальні додаткові втрати пов'язані з цим процесом.

Загальна схема роботи системи керування ресурсами зображена на рис. 9.

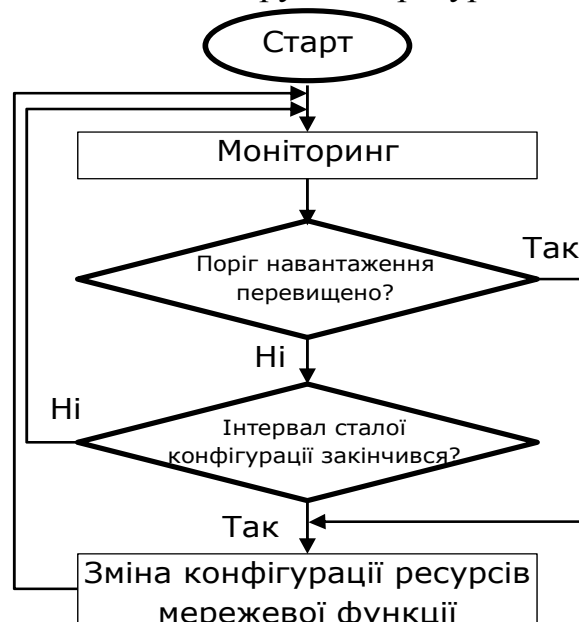


Рисунок 9 – Метод керування ресурсами системи

ВИСНОВКИ

У дисертаційній роботі вирішено важливе наукове завдання пов'язане з організацією обчислювальних ресурсів вузлів базової мобільної мережі та координації потоків між вузлами обслуговування у гібридному середовищі, яке складається із телекомунікаційної мережі та віртуальних обчислювальних сутностей розміщених у хмарі.

За підсумками вирішення завдання можна зробити наступні висновки:

1. Проаналізовано структуру телекомунікаційних сервісів, які надаються компаніями мобільного зв'язку, досліджено систему обслуговування службових потоків та підсистем базової мобільної мережі, виявлено, що збільшення службового навантаження призводить до збільшення необхідності у додаткових ресурсах, разом з тим нерівномірність завантаження вузлів інфраструктури призводить до їх простою, таким чином виникає потреба у впровадженні

технологій, які не призводять до простоїв обладнання, а також гарантують якість обслуговування навантаження протягом дня.

2. Визначено, що для задоволення ряду суперечливих вимог до мережевих функцій необхідно застосовувати гібридні системи, які дозволяють спільно використовувати технології апаратної та віртуальної реалізації мережевої функції.

3. Розроблено метод визначення місця розміщення та обсягів зарезервованих обчислювальних ресурсів віртуальних мережевих функцій у дата центрах оператора мобільного зв'язку, який гарантує якість надання телекомунікаційних сервісів з мінімально необхідними затратами ресурсів, виконуючи визначення їх достатньої конфігурації у гібридному середовищі доступних ресурсів, що дозволяє скоротити витрати на 13% у порівнянні з випадково обраною мономарою та на 47% у порівнянні з традиційним підходом до розгортання мережі.

4. Розроблено метод визначення розміру інтервалу часу сталої конфігурації обчислювальних ресурсів, який передбачає його змінну величину і врахування як витрат на реконфігурацію, так і коефіцієнту використання ресурсів, забезпечуючи гнучке використання ресурсів у віртуалізованому середовищі, зменшуючи відсоток вільних ресурсів на 42% у порівнянні з виділеним обладнанням і на 9% у порівнянні з існуючими аналогами та зменшуючи робоче навантаження у мережі.

5. Удосконалено розподілений метод локальної реконфігурації обчислювальних ресурсів віртуальної мережі у випадку відмови або перевантаження, який за рахунок децентралізованого керування та врахування міграційних витрат перерозподіляє віртуальні мережеві функції в нормальному та нештатному режимі із забезпеченням економічно обґрунтованого використання ресурсів, зменшуючи витрати в середньому на 21%.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Монографія

1. Математичні методи аналізу та керування телекомунікаційними мережами. Розділ 5 // Л.С. Глоба, О.М. Дяденко, А.Ю. Пилипенко, М.А. Скулиш. –К.: Інститут обдарованої дитини НАПН України. — ISBN 978-966-2633-84-9. — 2018. — С. 206-225.

Особистий внесок: система керування відображенням та плануванням ресурсів телекомунікаційної мережі з використанням технології віртуалізації.

Статті у наукових фахових виданнях

2. Скулиш М. А. Організація управління в багатоетапних системах масового обслуговування / М. А. Скулиш, С. В. Суліма // Вісник Національного університету «Львівська політехніка». Радіоелектроніка та телекомунікації. — 2015. — № 818. — С. 80–85.

Особистий внесок: аналітична модель для визначення кількості ресурсів, яку необхідно виділити для кожного етапу багатоетапної системи.

3. Skulysh M. Model for Efficient Allocation of Network Functions in Hybrid Environment / M. Skulysh, L. Globa, S. Sulima // Information and Telecommunication Sciences. — 2016. — № 1. — С. 39–45.

Особистий внесок: модель виділення ресурсів системи даних центрів для визначення оптимальної кількості ресурсів виділених мережевому функціональному вузлу.

4. Суліма С. В. Метод відновлення мережі у віртуалізованому середовищі / С. В. Суліма // Радіоелектроніка та інформатика. — 2017. — № 4. — С. 4–8.

5. Skulysh M. Hybrid resource management system for telecommunication network / M. Skulysh, S. Sulima // Advanced Information Systems. — 2018. — Vol. 2, № 1. — P. 47–51.

Особистий внесок: спосіб керування ресурсами телекомунікаційної мережі шляхом вибору оптимального інтервалу для перерозподілу системних ресурсів.

Статті у фахових виданнях України, які включені до міжнародних наукометричних баз

6. Суліма С. В. Система управління ресурсами в центрах обробки даних оператора мережі мобільного зв'язку / С. В. Суліма, М. А. Скулиш // Вісник НТУУ «КПІ». Серія Радіотехніка, Радіоапаратобудування. — 2017. — № 68. — С. 27-32. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: DOAJ, Web of Science, Index Copernicus, eLibrary.ru / РИНЦ та ін.)

Особистий внесок: метод динамічної конфігурації ресурсів.

7. Суліма С. В. Гібридна система управління ресурсами для віртуалізованих мережеских функцій / С. В. Суліма, М. А. Скулиш // Радіоелектроніка, інформатика, управління. — 2017. — № 1(40). — С. 16–23. (Наукове фахове видання України з технічних наук. Входить до міжнародних наукометричних баз: Web of Science, DOI, DOAJ, British Library, eLibrary.ru / РИНЦ, Index Copernicus та ін.)

Особистий внесок: метод побудови системи управління ресурсами для віртуальних мережеских функцій в центрах обробки даних оператора мережі мобільного зв'язку.

Статті у інших виданнях

8. Скулиш М. А. Метод управління ресурсами в датацентрах оператора мережі мобільного зв'язку / М. А. Скулиш, С. В. Суліма // Радіоелектроніка та інформатика. — 2015. — № 3. — С. 8–13.

Особистий внесок: модель для визначення оптимальної кількості датацентрів, топології їх розміщення та кількості ресурсів, які виділяються функціональним блокам системи, що розгортаються в дата центрах та передбачають використання декількох типів ресурсів.

Тези наукових доповідей

9. Skulysh M. Management of Multiple Stage Queuing Systems / M. Skulysh, S. Sulima // CADSM 2015 : 13-th International conference, 24–27 February 2015 : IEEE Digital Library. — Lviv–Polyana, 2015. — P. 431–433.

Особистий внесок: модель динамічного виділення ресурсів у багатоступінчатих системах.

10. Скулиш М. А. Організація управління в багатоступінчатих системах масового обслуговування / М. А. Скулиш, С. В. Суліма // Проблеми

телекомунікацій ПТ-2015 : 9-а міжнародна науково-технічна конференція, 21–24 квітня 2015 : матеріали конференції. — Київ, 2015. — С. 319–321.

Особистий внесок: метод динамічного керування ресурсами багатоступінчастих систем.

11. Скулиш М. А. Управління ресурсами в віртуалізованих мережах оператора мобільного зв'язку / М. А. Скулиш, С. В. Суліма // Інфокомунікації – сучасність та майбутнє : 5-а міжнародна науково-практична конференція, 29-30 жовтня 2015 : матеріали конференції. — Одеса, 2015. — С.84–88.

Особистий внесок: метод початкової організації гетерогенних ресурсів мережі.

12. Globa L. Method for resource allocation of virtualized network functions in hybrid environment / L. Globa, M. Skulysh, S. Sulima // 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 6–9 June 2016 : IEEE Digital Library. — Varna, Bulgaria, 2016. — P. 1–5.

Особистий внесок: метод адаптивної конфігурації гетерогенних ресурсів мережі.

13. Суліма С. В. Алгоритм відображення та планування віртуалізованих функцій в мережі мобільного зв'язку / С. В. Суліма, М. А. Скулиш // Проблеми телекомунікацій ПТ-2016 : 10-а міжнародна науково-технічна конференція, 19–22 квітня 2016 : матеріали конференції. — Київ, 2016. — С. 372–374.

Особистий внесок: метод відображення та планування віртуалізованих мережевих функцій.

14. Skulysh M. Service deployment aspects in the systems with network function virtualization / M. Skulysh, S. Sulima // Radio Electronics & Info Communications : First International Conference, 11–16 September 2016 : IEEE Digital Library. — Kyiv, 2016. — P. 1–7.

Особистий внесок: підхід до моделювання та управління розподілом ресурсів віртуалізованих мережевих функцій у мережі телекомунікаційного оператора.

15. Sunduchkov K. Sampling for direct search method of all system implementations developed according to the requirements of the technical design specification / K. Sunduchkov, M. Skulysh, S. Sulima, B. Savchuk // 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), 10–13 October 2017 : IEEE Digital Library. — Kharkov, 2017. — P. 83–86.

Особистий внесок: формалізація модифікованого методу оптимізації побудови системи зв'язку.

16. Суліма С. В. Метод відновлення вузла у мережах NFV / С. В. Суліма // Проблеми телекомунікацій ПТ-2017 : 11-а міжнародна науково-технічна конференція, 18–21 квітня 2017 : матеріали конференції. — Київ, 2017. — С. 270–272.

17. Sulima S.V. Model for allocation of network slices in heterogeneous environment / S. V. Sulima // Modern challenges in telecommunications : Twelfth International Scientific Conference, 16–20 April 2018 : conference proceedings. — Kyiv, 2018. — С. 42–44.

АНОТАЦІЯ

Суліма С.В. Методи реконфігурації обчислювальних ресурсів базової мережі на основі технології віртуалізації. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.12.02 - телекомунікаційні системи та мережі. – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» МОН України, Київ, 2019.

У дисертаційній роботі вирішено важливу науково-технічну задачу, яка пов'язана з розробкою наукових основ організації обчислювальних ресурсів вузлів обслуговування та потоків між ними у гібридному середовищі, яке складається із апаратних засобів та віртуальних сутностей, з метою підвищення коефіцієнту використання обчислювальних ресурсів за рахунок еластичності мережевих функцій, що дозволяє гарантувати допустимий час обслуговування трафіку сигналізації та контролювати витрати обчислювальних ресурсів в умовах зростання обсягу трафіку та кількості сервісів.

Запропоновано аналітичний підхід до задачі розміщення віртуальних мережевих функцій та резервування необхідних для їх належної роботи обчислювальних ресурсів, вирішено задачу організації керування обчислювальними ресурсами мережевих функцій протягом дня шляхом визначення інтервалів часу сталої конфігурації ресурсів мережевих функцій та коригуванням їх відповідно до поточного прогнозу навантаження, вирішується задача відновлення віртуалізованої мережі, що постраждала від відмови на нижчорозташованому фізичному вузлі або у випадку перевантаження, за рахунок міграційних процесів, що виконуються децентралізованими вузлами керування.

Ключові слова: мобільна мережа, базова мережа, віртуалізація, NFV, система керування навантаженням, конфігурація ресурсів, реконфігурація мережі.

АННОТАЦИЯ

Сулима С.В. Методы реконфигурации вычислительных ресурсов базовой сети на основе технологии виртуализации. – На правах рукописи.

Диссертация соискание ученой степени кандидата технических наук по специальности 05.12.02 - телекоммуникационные системы и сети. – Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского» МОН Украины, Киев, 2019.

Анализ текущей ситуации в области беспроводной связи показывает увеличение служебной нагрузки, что приводит к увеличению необходимости в дополнительных вычислительных ресурсах, вместе с тем неравномерность загрузки узлов инфраструктуры приводит к их простоям, таким образом возникает потребность во внедрении технологий, которые уменьшают простои оборудования и гарантируют качество обслуживания нагрузки на заданном уровне в течение дня. Проведенные исследования показали, что наиболее перспективным направлением развития телекоммуникационной отрасли является преобразование сетевых архитектур путем внедрения сетевых функций в программном обеспечении, которое работает на стандартной аппаратной платформе, позволяя

упростить процессы масштабирования, реконфигурации ресурсов и развертывания сервисов.

Чтобы справиться со значительным ростом трафика служебных данных мобильный оператор осваивает технологии виртуализации сети и облачных вычислений, чтобы построить масштабируемые и гибкие мобильные сети. В отличие от существующей статической архитектуры базовой сети LTE предлагается система, в которой выделены аппаратные сетевые функции обслуживают часть служебной нагрузки, а часть направляется на обработку с использованием виртуализированных сетевых функций в арендованных облаках дата центров.

В диссертационной работе предложено решение по управлению конфигурацией ресурсов в сетях оператора с гибридной структурой, которая характеризуется наличием как выделенных специализированных аппаратных ресурсов, так и гибких виртуализированных вычислительных ресурсов.

В первом разделе диссертации выполнен обзор литературы по теме исследования, который позволил выявить существующие проблемы в организации вычислительных ресурсов сети.

Во втором разделе предложено аналитический подход к моделированию и конфигурации вычислительных ресурсов виртуализированных сетевых функций в базовой сети телекоммуникационного оператора с помощью которого осуществляется первичное резервирования ресурсов сетевых функций перед началом функционирования системы. Определив целевую функцию задачи «встраивания» виртуальных сетевых функций на физическую инфраструктуру дата центров и распределения ресурсов и учитывая доступные ресурсы на узлах и их стоимость, интенсивность избыточной служебной нагрузки и допустимые задержки, строится метод определения топологии выделения ресурсов и их необходимого зарезервированного объема. Согласно решению, каждой сетевой функции резервируется определенное количество ресурсов на основе оценки ее наибольшей потребности в них; мгновенные потребности различных сетевых функций динамично удовлетворяются путем активации необходимой конфигурации виртуальных машин при выполнении.

В третьем разделе решена задача построения системы управления вычислительными ресурсами виртуализированных сетевых функций для предоставления гибкости системе обслуживания. Предложен метод определения величины интервала времени постоянной конфигурации ресурсов сетевых функций, который обеспечивает динамическую настройку работы системы, позволяющий повысить коэффициент использования ресурсов и разгрузить сетевые узлы, обеспечивая приемлемое качество обслуживания служебных потоков. Представлен подход к прогнозированию нагрузки, учитывающий как накопленные за длительный срок данные статистики, так и последние тенденции, наблюдаемые в сети, что позволяет достигать оптимального соотношения затрат на управление и итогового значения качества обслуживания.

В четвертом разделе решается проблема восстановления виртуализированных сети, пострадавшей от отказа в нижележащем физическом узле или в случае перегрузки. Предлагаемый подход выполняется распределенным способом. В

исследовании решается проблема расположения децентрализованных узлов управления с учетом параметров задержки и устойчивости, предлагается целевая функция комбинаторной оптимизационной задачи расположения узлов управления. Предложен метод локальной реконфигурации узлов после перегрузки или отказа, который использует эвристическую модель для осуществления миграции виртуализированных сетевых функций.

Проведены эксперименты по исследованию предложенных методов, которые доказали их целесообразность.

Ключевые слова: мобильная сеть, базовая сеть, виртуализация, NFV, система управления нагрузкой, конфигурация ресурсов, реконфигурация сети.

ABSTRACT

Sulima S.V. Reconfiguration methods of the computing resources for the core network based on virtualization technology. – On the rights of the manuscript.

A thesis submitted in fulfilment of the Candidate of Sciences degree in technical sciences on specialty 05.12.02 - telecommunication systems and networks. – National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» of Ministry of Education and Science of Ukraine, Kyiv, 2019.

The thesis has solved an important scientific and technical task related to the development of scientific bases for the computing resources organizing for service nodes and flows between them in a hybrid environment, which consists of hardware and virtual entities, in order to increase the utilization of computational resources due to elasticity of network functions, which allows guaranteeing the permissible time for signaling traffic servicing and controlling the costs of computing resources in conditions of growth in traffic volume and number of services.

The analytical approach to the task of the allocation of virtual network functions and the reservation of the computational resources necessary for their proper work is proposed, the task of managing the computer resources of the network functions during the day is solved by determining the time intervals of the constant configuration of the network functions resources and adjusting them in accordance with the current load forecast, a virtualized network with failure on underlying physical node or overload recovery task is considered by means of migration performed by decentralized management nodes.

Keywords: mobile network, core network, virtualization, NFV, traffic management system, resource configuration, network reconfiguration.